Supplementary Figure 1. Study Design.

We designed this study as a disease landscape study with no pre-specified hypothesis. For biologically and clinically relevant sub-group detection, we profiled a final number of tumors n=300 higher than any of the studies in public domain (April 2010) from any single group. We generated two types of genomics data a) Whole genome sequencing of 49 GC tumors and recurrent mutation information (Wong et al, Nat Comm 2014) that informed selection of genes for targeted re-sequencing done in this study b) Gene expression, Copy copy number profiling and targeted re-sequencing on an additional n=251 cases.

Three hundred primary independent GC specimens were procured at the time of total or subtotal gastrectomy at Samsung Medical Centre, Seoul, Korea, from 2004 to 2007, and frozen at -80°C. The protocol was approved by Samsung Medical Centre Institutional Review Board (IRB No. 2010-12-088). We obtained the informed consent according to the IRB protocol. We clinically annotated the tumors but de-linked them from personally identifiable information. This cohort has a median follow up time of 86.4 months (Range: 53.1 – 106.6 months). We procured all of the tissue specimens at chemo-naïve state during primary resection of gastric cancer. No subjects received neoadjuvant chemotherapy or chemoradiation therapy. Ninety eight subjects received post-operative chemotherapy or chemoradiation therapy (CRT) in this cohort. The post-operative surveillance program for recurrence is to follow up every 6 months until 5-years from the date of surgery.

We selected cases based on the following criteria: histologically confirmed adenocarcinoma of the stomach; surgical resection of primary GC; age ≥18 years; complete pathological, surgical, treatment, and follow-up data. Two expert gastrointestinal pathologists (K.M.K, I.D.) reviewed hematoxylin and eosin stained slides to select cases with estimated carcinoma content of at least 60%.
Molecular characterization (N=300)

SMC Biobank
GC patients with primary cancer resection (with adequate tissue specimens and clinical annotation) (2004-2007)

ACRG cohort (N=300)

Gene expression profiling, CNV, Targeted sequencing

4 molecular classes identified

TCGA Validation cohort (N=184)

Target sequencing list derived

Singapore Validation cohort (N=277)

TCGA Validation cohort (N=184)

Independent validation cohorts

WGS (N=49)

Molecular characterization (N=300)

ACRG cohort (N=300)
Supplementary Figure 2. Principal component analysis of gene expression data

Signatures of MSI/MSS, Proliferation, and mesenchymal exist in gastric carcinoma and strongly correlate with the main Principal components (PC) of data: PC1 correlates with EMT and anticorrelates with Proliferation, PC3 correlates with MSI/MSS and cytokine signature, and PC2 correlates with stomach tissue signature. In the MSS/EMT- tumors, proliferation typically correlates with TP53 mutation and CIN.

Nature Medicine: doi:10.1038/nm.3850
Supplementary Figure 3. Distribution tails of EMT signature and MSI signature define samples in the Mesenchymal and MSI subtypes.

The signature thresholds as shown in the plot below were defined using qqplots of the respective distributions. MSI and EMT outliers are mutually exclusive (Fisher P<6e-6). MLH1 methylation was available for this set as MSI marker (red dots) and shows a strong association with MSI signature.
Supplementary Figure 4. TP53 mutation status in non-EMT non-MSI group.
We considered a two-gene TP53 signature (average of MDM2 and CDKN1A expression), and built the ROC-AUC curve of TP53 signature vs TP53 mutations. We assigned the signature threshold at the Youden index of the ROC-AUC curve and assigned group membership into TP53-active and inactive based on this threshold.

Note: A caveat of using TP53 transcriptional signature as a translational tool is that samples with amplified MDM2 that overexpress MDM2 are classified as TP53+, although the TP53 pathway is inactive in these samples. Given the relatively small number of such samples (~2%), this misclassification is unlikely to impact other analysis.
Supplementary Figure 5. Focal copy number events and relation with gene expression

Focal copy number events for CCND1, CCNE1, EGFR, ERBB2, FGFR2, GATA4, MDM2, MYC associate with significantly increased gene expression.
Supplementary Figure 6. IHC assays for EGFR and ERBB2 significantly associate with focal amplifications and gene expression.

We illustrate here the correlation between mRNA expression (X-Axis), CNV (Y-Axis) and IHC (Overlay, red). Gene expression and CNV data are described in the methods section. The IHC data generation is described here. IHC data: EGFR:: we used the anti-NCL-L-EGFR-384 mouse monoclonal primary antibody (Novocastra, UK; 1:100 dilution). ERBB2:: we used the PATHWAY® HER2 (4B5) rabbit monoclonal antibody (Ventana Medical Systems, No dilution required). We used Ventana BenchMark XT automated slide processing system according to the manufacturer’s protocol.
Supplementary Figure 7. Correlative analysis of ACRG signatures with expression signatures from TCGA Gastric (Nature 2014) and Lei et al (Gastroenterology, 2013).

EMT.ACRG, MES.SG and C1.TCGA predominantly correlate with each other and thus appear to track the same patient group consistently. The MSI (ACRG) signatures correlate with C3.TCGA and PRO.SG (Proliferation) as well as weakly with Cytokine and C2.TCGA signatures and C2.TCGA and MET.SG (Metabolic). The TP53 (ACRG) signature likely tracks patients that are not exclusively related to any signatures by SG and TCGA. Signatures defining expression subtypes in the TCGA Gastric, ACRG Gastric and Singapore Gastric datasets are shown in the legend.

TCGA Signatures C1.TCGA, C2.TCGA, C3.TCGA and C4.TCGA comprise of 10 genes each. MES.Lei, PRO.Lei and MET.Lei comprise of 200, 200 and 186 genes respectively.
Supplementary Figure 8. Cancer Cell Line Encyclopedia (CCLE) cell line data analysis

We used the gene expression signatures to study whether the tumor subtypes are proportionally represented in cell lines. Using the Cancer Cell Line Encyclopedia (CCLE), gastric and esophagus cancer cell line panel, we calculated the hypermutation and copy number variation across a set of 63 cell lines and the gene expression signatures for the mesenchymal, MSI and TP53 status. We observed a significant association of RNA and DNA profiles consistent with the association in primary tumor samples (P <1e-3), but the MSS/TP53+ subtype was highly under-represented, illustrative of the lack of representation of early TP53-wt tumors in cell line panels.
Supplementary Figure 9. Gene expression molecular subtypes and their association with DNA features are evaluated in TCGA Colorectal cancer dataset.
Supplementary Figure 10. MSI Assays – Concordance between immunohistochemistry and Pentaplex PCR assay.

(A) MLH1 immunohistochemistry showing loss of expression in gastric cancers. (B) Pentaplex PCR results in a case with MSI-high showing allelic size variations in all NR27, NR21, BAT25, BAT26, and NR24 mononucleotide repeat markers. (C) MLH immunohistochemistry showing preserved expression of MLH1 protein in tumors. (D) Pentaplex PCR in a case with MSS showing allelic size variations in none of the markers.
### Supplementary Table 1: Multivariable analysis of overall survival (ACRG Cohort)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio</th>
<th>95% CI for hazard ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtypes (MSS/TP53+, MSS/TP53- and MSI vs MSS/EMT)</td>
<td>1.899</td>
<td>(1.109 – 3.254)</td>
<td>0.0195</td>
</tr>
<tr>
<td>Age (62 &lt; vs 62&gt;=)</td>
<td>2.014</td>
<td>(1.244 – 3.262)</td>
<td>0.0044</td>
</tr>
<tr>
<td>Sex (female vs male)</td>
<td>1.187</td>
<td>(0.727 – 1.937)</td>
<td>0.4926</td>
</tr>
<tr>
<td>Location (non-antrum vs antrum)</td>
<td>0.652</td>
<td>(0.413 – 1.031)</td>
<td>0.0674</td>
</tr>
<tr>
<td>WHO (signet/PD/others vs W/D ~M/D tubular)</td>
<td>1.385</td>
<td>(0.612 – 3.133)</td>
<td>0.4348</td>
</tr>
<tr>
<td>Lauren (diffuse vs intestinal)</td>
<td>0.619</td>
<td>(0.272 – 1.405)</td>
<td>0.2512</td>
</tr>
<tr>
<td>AJCC stage(Ib/II vs III/ IV)</td>
<td>2.932</td>
<td>(1.750 – 4.912)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>EBV (negative vs positive)</td>
<td>0.625</td>
<td>(0.258 – 1.510)</td>
<td>0.2962</td>
</tr>
<tr>
<td>Lymphovascular invasion(negative vs positive)</td>
<td>1.442</td>
<td>(0.834 – 2.492)</td>
<td>0.1901</td>
</tr>
<tr>
<td>Perineural invasion(negative vs positive)</td>
<td>1.715</td>
<td>(1.094 – 2.686)</td>
<td>0.0186</td>
</tr>
</tbody>
</table>